VNU Journal of Science: Natural Sciences and Technology
Journal homepage: https://js.vnu.edu.vn/NST

Original Article

In silico Identification of the HLA-DRB1*15:01 Restricted B-cell and T-cell Epitopes of the Autoantigens MOG and AQP4 Associated with Autoimmune Neuropathy

Le Thi Thanh Nhan[1] , Ha Thao Van[1], Bui Phuong Thao[1], Le Lan Phuong[1], Trinh Hong Thai[1,2,*]

[1]*VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam*
[2]*Institute of Biomedical Science and Technology, 67-116 Nguyen Xien, Thanh Liet, Hanoi, Vietnam*

Received 10th July 2025
Revised 20th January 2026; Accepted 23rd January 2026

Abstract: Immunogenicity plays a crucial role in determining the pathogenesis of autoimmune neuropathy associated with MOG and AQP4 proteins. In this study, we used computational tools, including MHC-II Binding Prediction and B Cell Epitope Prediction tools from IEDB, Clustal Omega, SWISS-MODEL, AbodyBuilder, HDOCK, AlphaFold Protein Structure Database, and Protein Data Bank, to identify potential HLA-DRB1*15:01 restricted T-cell and B-cell epitopes of the autoantigens MOG and AQP4. Subsequently, the peptide sequences of bacterial proteins identical to the epitopes of MOG and AQP4 proteins were analyzed using BLASTp, MoRFchibi SYSTEM, IUPred3, SEG, miPepBase, and NCBI databases to detect the presence of MoRF, SliM, and LCR. Our findings revealed that $MOG_{22\text{-}LQVSSSY\text{-}28}$, $MOG_{80\text{-}YRNGKD\text{-}85}$, $MOG_{179\text{-}QYRLRGKL\text{-}186}$ and $AQP4_{75\text{-}CTRKIS\text{-}80}$ were common sequences of the HLA-DRB1*15:01 restricted T-cell and B-cell epitopes of the MOG and AQP4 autoantigens, as well as MoRFs, SLiMs, and LCRs present in bacterial peptides identical to four core epitopes associated with autoimmune diseases. This study has provided information to further understand the pathogenesis of autoimmune diseases associated with MOG and AQP4.

Keywords: Epitope, MOG, AQP4, HLA-DRB1*15:01, autoimmune neuropathy.

1. Introduction

Myelin oligodendrocyte glycoprotein (MOG) and Aquaporin-4 (AQP4) are two autoantigens implicated in autoimmune neurological disorders, including multiple sclerosis (MS), neuromyelitis optica spectrum disorder (NMOSD), and acute disseminated encephalomyelitis (ADEM) [1]. The specific roles of MOG and AQP4 in the pathogenesis of these diseases remain incompletely understood [2].

Animal model studies and genetic association analyses have identified major histocompatibility complex (MHC) class II

* Corresponding author.
E-mail address: thaithvnu@yahoo.com

https://doi.org/10.25073/2588-1140/vnunst.5798

molecules, particularly alleles of the human leukocyte antigen (HLA)-DR isotype, as genetic susceptibility factors in MOG- and AQP4-associated autoimmune conditions [3-5]. HLA-DRB1*15:01 has been recognized as a potential risk allele for ADEM, yet appears to have a protective role in NMOSD [6, 7]. Notably, the frequency of this allele was significantly lower in NMOSD patients than in MS patients (P = 0.0001), suggesting a potential protective effect against the development of MS in NMOSD individuals [3]. In 2010, Marcelo et al. further confirmed the protective association of the HLA-DRB1*15:01 allele with both MS and NMOSD [4].

On the other hand, MS, ADEM and NMO are autoimmune diseases triggered by dysregulated immune responses, particularly involving T-cells and B-cells reacting against self-antigens in the central nervous system [1]. An epitope is a specific molecular region on an antigen recognized by the immune system, typically by T or B lymphocytes [5]. T-cell epitopes are peptide fragments processed from antigenic proteins, presented by MHC molecules on antigen-presenting cells, and recognized by T-cell receptors [5]. In contrast, B-cell epitopes are surface-exposed amino acid sequences or structural motifs on antigens that are directly recognized by antibodies [5]. To date, no published studies have explored the interaction of T-cell and B-cell epitopes with the HLA-DRB1*15:01 molecule in the context of MS, ADEM, or NMOSD using *in silico* approaches. Therefore, this study aims to predict and characterize T-cells and B-cells epitopes associated with MOG and AQP4 and evaluate their interactions with HLA-DRB1*15:01, thereby contributing to a better understanding of immune mechanisms involved in these autoimmune diseases.

In addition to epitope prediction, this study also identifies sequence similarities between human epitopes and viral/bacterial proteins, and investigate the potential functional implications of these similarities-particularly within intrinsically disordered regions. Previous research has shown that certain pathogens carry peptide sequences that are identical or structurally similar to human proteins, a phenomenon that can lead to molecular mimicry and cross-reactivity, ultimately triggering autoimmune responses [6].

Intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) lack stable tertiary structure but are functionally important due to their conformational flexibility, allowing interaction with multiple partners and involvement in signaling and regulatory networks [7]. These regions frequently contain molecular recognition features (MoRFs), short linear motifs (SLiMs), and low complexity regions (LCRs) reported to be commonly present in IDPs/IDPRs-all of which are known to play critical roles in cellular processes.

MoRFs are peptide sequences of 10–70 amino acids that involved in protein-protein interactions. IDPs can simultaneously recruit multiple MoRFs during interactions [8]. MoRF-containing proteins are abundantly present in ribosomes, nuclei, and microtubules and are involved in translation, protein transport, protein folding, and DNA interactions [11, 12]. SLiMs, consisting of 3-10 amino acid residues, are versatile interaction modules that regulate signaling, subcellular localization, post-translational modifications, and protein stability [11, 13]. Many pathogenic viruses and bacteria have evolved to incorporate host-mimicking SLiMs, thereby hijacking host cellular processes [9]. In addition, LCRs are also involved in important functions such as regulating protein-protein interactions, nucleic acid-protein interactions, subcellular protein localization, antigen processing and diversification [7].

Collectively, MoRFs, SLiMs, and LCRs facilitate diverse and critical protein interactions. However, their presence in microbial peptides identical to human MOG/AQP4 epitopes and presented by HLA-DRB1*15:01 has not yet been systematically investigated. Therefore, in this study, we conducted a computational analysis to assess the occurrence of MoRFs, SLiMs, and LCRs in

viral and bacterial sequences identical to predicted MOG/AQP4 epitopes, aiming to elucidate the potential mechanisms of cross-reactivity and autoimmune pathogenesis.

2. Experimental

2.1. Sequence and Structure of MOG and AQP4

The FASTA sequences of MOG (AlphaFold DB ID: AF_AFQ16653F1) and AQP4 (PDB ID: 3GD8) were obtained from AlphaFold DB (https://alphafold.ebi.ac.uk/) and Protein Data Bank (https://www.rcsb.org/), respectively. These sequences represent the wild-type (normal human) forms of the proteins, which have been experimentally determined. The MOG structure from AlphaFold and the AQP4 structure from the Protein Data Bank are reliable sources commonly used as reference standards in molecular modeling and docking studies [10], [11]. These sequences are not specific to patients with MS, NMO, or ADEM, but rather reflect the typical sequences found in the general human population. In most cases, the MOG and AQP4 sequences are identical between patients and healthy individuals, except for rare mutations not commonly associated with these diseases [12-14]. Therefore, using the reference sequences ensures representativeness, reproducibility, and comparability in research, while avoiding bias caused by individual genetic variations.

2.2. Methods

2.2.1. Prediction of $CD4^+$ T-cell Epitope

MHC-II Binding Predictions tool from IEBD Analysis Resource was used to predict MHC class II $CD4^+$ T-cell epitopes based on Inhibitory Concentration (IC50) values for peptides binding to specific HLA molecules (http://tools. iedb.org/mhcii/). In this study, HLA-DRB1*15:01 was selected for the interaction analysis with antigenic peptides. All the potential epitopes predicted to bind to HLA-DRB1*15:01 with IC50 score < 500 nM showing stronger than average binding affinity were selected for further study.

2.2.2. Prediction of B-cell Epitopes

Sequence-based: The IEBD Analysis Resource was used for analysis, with particular focus on two autoantigens, MOG and AQP4. These sequences were then subjected to various prediction methods, including Bepipred linear epitope prediction, Emini surface accessibility, Kolaskar and Tongaonkar antigenicity, Parker hydrophilicity, Chou and Fasman beta turn, and Karplus and Schulz flexibility prediction, all of which can be found on the B Cell Epitope Prediction Tools website (http://tools.iedb.org/bcell/). These methods aim to predict the likelihood that specific regions within the protein bind to the B-cell receptor, are present on the surface, are immunogenic, are located in a hydrophilic region, and are part of the beta turn region.

Structure-based: ElliPro (http://tools.iedb.org/ellipro/) was used to predict B-cell epitopes. ElliPro can predict both linear and conformational epitopes based on the 3D structure of the protein.

For each method/tool, each epitope has 7 amino acids and thresholds are set according to their default values. The predicted results are a graph in which regions with scores higher than the thresholds are more likely to be B-cell epitopes. All graphs provided by the tools were compared. Epitopes predicted with high confidence that fell above the threshold were selected for further analysis.

2.2.3. Screening for Potential Core Epitopes Shared between $CD4^+$ T-cell and B-cell

To find the final predicted epitopes, all the resulting T and B-cell epitopes were clustered using the Clustal Omega tool (//www.ebi.ac.uk/Tools/msa/clustalo/) and then compared to each other to find the common core epitope.

2.2.4. Receptor Preparation

3D model construction: The 3D model of the HLA-DRB1*15:01 molecule is not available in the Protein Data Bank (https://www.rcsb.org/), so SWISS-MODEL (https://swissmodel.expasy.org/) was used to model the structure of HLA-DRB1*15:01 based on the HLA-DRB1*15:01 protein sequence (IMGT HLApro: HLA00865) collected from

the IPD-IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/), the alpha chain sequence of the HLA-DR complex (ID: P01903) was obtained from the UniProt database (http://www.uniprot.org/).

Identification of the antigen-binding region: Anti-MOG and anti-AQP4 antibodies are usually IgG1 [12, 15]. At the time of this study, no specific structural data were available for human IgG or IgG1. Therefore, to investigate the complement-binding regions of human IgG antibodies, we used the IgG Fab structure from the Protein Data Bank (PDB ID: 5JQD) and the AbodyBuilder tool (https://opig.stats.ox.ac.uk/webapps/sabdab-abpred/sabpred/abodybuilder/). PDB ID 5JQD was selected as it provides a high-resolution (2.20 Å) representation of the Fab region, fully capturing both heavy and light chains along with detailed atomic interactions. This structural clarity facilitates the understanding of how Fab regions recognize and bind to epitopes. Additionally, 5JQD exhibits typical immunoglobulin folding, making it an ideal model for structural comparisons and research applications [16].

Before the interaction analysis, all molecular structures were prepared and processed for docking using the Dock Prep tool (UCSF Chimera X) [17].

2.2.5. Protein-peptide Molecular Docking

The HDOCK program was used to predict the binding structures between the MOG/AQP4 epitopes and HLA-DRB1*15:01 and the IgG Fab region (http://huanglab.phys.hust.edu.cn/hdock/). The structural model of the bound complex was selected with the best binding score and the correct docking of the peptide/antigen binding region of the protein structure.

2.2.6. Identification of Peptides Identical to MOG and AQP4 Epitopes on Viruses and Bacteria Associated with Autoimmune Neuropathy

miPepBase Database: Information on viral and bacterial peptides related to core epitopes was obtained from miPepBase - a database of experimentally identified peptides associated with autoimmune disease (Mimicry Peptide Database, http:// proteininformatics.org/mkumar/mipepbase/search.html). Sequence comparisons were performed using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

NCBI database: The BLAST was used to search for regions of sequence similarity between the potential epitopes and viral (ID taxids: 10239) and bacterial (ID taxids: 2) protein sequences in the NCBI non-redundant (nr) protein database. All parameters selected as default settings except word size = 2. Search results with query coverage and sequence identity ≥ 80% were selected as criteria for similarity. Identical peptides in viruses and bacteria associated with autoimmune diseases were selected for further analysis.

2.2.7. Investigating the Disordered Regions (MoRFs, SLiMs and LCRs) Present in Identical Peptides

The presence of MoRFs in identical peptides was analyzed using MoRFchibi SYSTEM (MoRFCHiBi_Web mode, https://morf.msl.ubc.ca/) due to its ability to provide highly accurate predictions. For SLiM prediction in mimitopes, IUPred3 (ANCHOR2 analysis type selection, https://iupred3.elte.hu/) was used and SEG (https://mendel.imp.ac.at/METHODS/seg.server.html) was used to find LCRs. MoRFCHiBi_Web, ANCHOR, ANCHOR2 and SEG were used at default parameters. MoRFs/SLiMs/LCRs were considered to be present in the host or pathogen mimitopes if at least half of the amino acids of the mimitopes overlapped with these regions.

3. Results and Discussion

3.1. Potential Linear Epitopes of MOG/AQP4 Shared between CD4$^+$ T-cell and B-cell

The presentation of antigenic peptides via HLA class II molecules triggers the activation of helper T-cells and the production of autoantibodies [18]. Therefore, it is crucial to consider the relationship between T-cell and B-cell epitopes in order to select epitopes that can induce both humoral and cell-mediated immune responses. Our research has identified 9 CD4$^+$ T-cell core epitopes from MOG and 8

$CD4^+$ T-cell core epitopes from AQP4, as well as 9 constitutive B-cell core epitopes from MOG and 7 constitutive B-cell core epitopes from AQP4. Upon comparing the T-cell and B-cell core epitope sequences obtained through clustering using the Clustal Omega tool, we found that 4 sequences from B-cell were shorter than those from T-cells. These 4 core epitopes, including $MOG_{22\text{-LQVSSSY-28}}$, $MOG_{80\text{-YRNGKD-85}}$, $MOG_{179\text{-QYRLRGKL-186}}$, and $AQP4_{75\text{-CTRKIS-80}}$ (highlighted bold in Table 1), were likely processed through antigen presentation.

Table 1. Linear core epitopes of MOG/AQP4 shared between $CD4^+$ T-cell and B-cell

Autoantigen	Peptide			
	Cell	Position	Length	Sequence
MOG	T	21-30	10	LLQVSSSYAG
	B	22-28	7	LQVSSSY
$MOG_{22\text{-LQVSSSY-28}}$		22-28	7	LQVSSSY
MOG	T	69-85	17	YRPPFSRVVHLYRNGKD
	B	80-97	18	YRNGKDQDGDQAPEYRGR
$MOG_{80\text{-YRNGKD-85}}$		80-85	6	YRNGKD
MOG	T	174-186	13	IFLCLQYRLRGKL
	B	179-187	9	QYRLRGKLR
$MOG_{179\text{-QYRLRGKL-186}}$		179-186	8	QYRLRGKL
AQP4	T	71-93	23	VAMVCTRKISIAKSVFYIAAQCL
	B	75-80	6	CTRKIS
$AQP4_{75\text{-CTRKIS-80}}$		75-80	6	CTRKIS

3.2. In silico Interactions of Epitope-HLA and Epitope-Fab

3D structural model of HLA-DRB1*15:01: The 3D structure of HLA-DRB1*15:01 was constructed using SWISS-MODEL (Figure 1A). The resulting model shares a sequence similarity of 96.67% with the crystal structure model of HLA-DR1 CLIP102-120 (SMTL ID: 3pdo.1) (Figure 1B). The quality of the structure was evaluated using SWISS-MODEL's structure evaluation metrics, with a GMQE of 0.69 and a QMEANDisCo Global of 0.85±0.05, indicating a reliable model. The Ramachandran plot showed that 97.57% of the amino acids are located in the favorable region (Figure 1C). When compared to similar sized experimental structures, the modeled HLA-DRB1*15:01 structure had a normalized QMEAN score within |Z-score| < 1 of the Non-redundant set of PDB structures (Figure 1D), further supporting the quality of the model.

The 3D structure predicted by SWISS-MODEL was compared to the template 3D structure (SMTL ID: 3pdo.1) using ChimeraX software to analyze any differences between them. The alignment results showed a high level of similarity in both sequence and 3D shape (Squence alignment score = 966.8), with a very small Root Mean Square Deviation (RMSD = 0.068 Å, across all 184 pairs) (Figure 1E). However, there are still some structural differences in the B chain, especially the majority of these amino acids are located in the protein/peptide interaction region of MOG, AQP4 (Figure 1B).

Antigen-binding region of human IgG Fab structure: The analysis using ABodyBuilder showed that the newly generated antibody model shared 93% similarity with the heavy chain and 97% similarity with the light chain of the reference IgG Fab structure. The model exhibited a root-mean-square deviation (RMSD) of approximately 1.45 Å in the complementarity-determining regions (CDRs) when compared to the reference structure, indicating a high-quality structural prediction in the antigen-binding regions. This is further illustrated in Figure 2, which highlights the CDR region.

Molecular interactions: To evaluate the interaction between the predicted potential

epitope with HLA and autoantibodies, in this study, HLA-DRB1*15:01 and Fab of human IgG were treated as receptors in the molecular docking process with epitopes (ligands) using the HDOCK program. The results showed that all four epitopes bound at the peptide/antigen binding site (the binding site is shown in Figure 3 with $MOG_{179\text{-}QYRLRGKL\text{-}186}$ as the representative epitope). The selected model had high reliability, with the lowest docking score, and a confidence score > 0.5, indicating that the two molecules could bind to each other (Table 2).

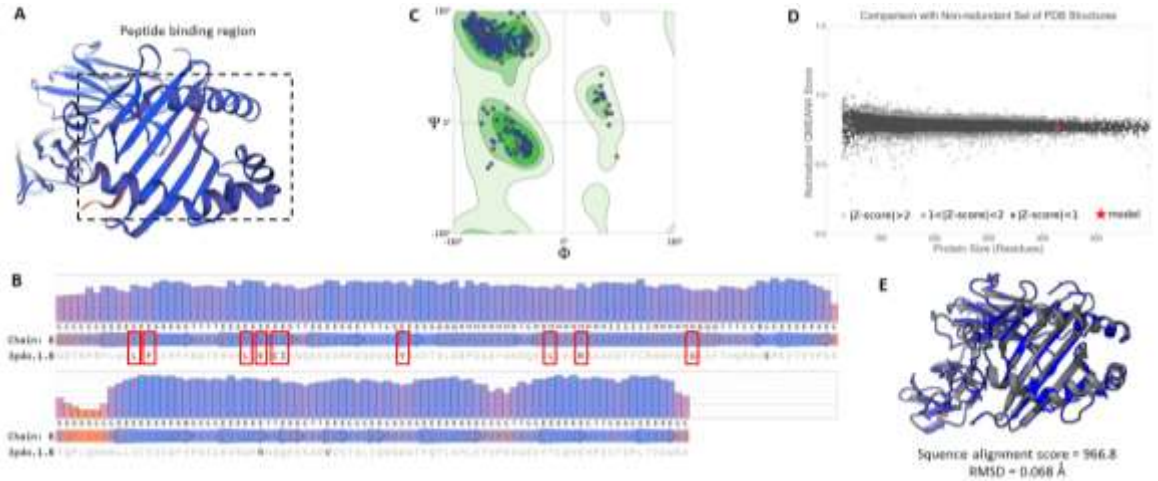

Figure 1. Evaluation of the HLA-DRB1*15:01 model.

*(A: Structural model HLA-DRB1*15:01; B: Sequence differences (chain B) of target (HLA-DRB1*15:01) and template (SMTL ID: 3pdo.1) (The red squares indicate the amino acids that differ in the antigen-binding region between the two models); C: Ramachandran plot (the white area was the outlier, where there were almost no Φ/Ψ pairs; the light green area was the area created by the first contour line, where there were 99.7% of the Φ/Ψ pairs; the green area was the area created by the 2nd contour line, where there were 95.0% Φ/Ψ pairs, and the dark green area was the area of interest, created by the 3rd contour line, where there were 80.0% Φ/Ψ pairs); D: Comparison with a non-redundant set of PDB structure (The x-axis shows protein length, represented by the number of residues). The y-axis was the normalized QMEAN score. Each dot represented an experimental protein structure, and the |Z-score| indicated the standard deviation from the mean. HLA-DRB1*15:01 model is represented as a red star; E: Superposition of two structures HLA-DRB1*15:01 (blue) and the structural model of HLA-DR1 CLIP102-120 (SMTL ID: 3pdo.1, gray) using the Needleman–Wunsch algorithm).*

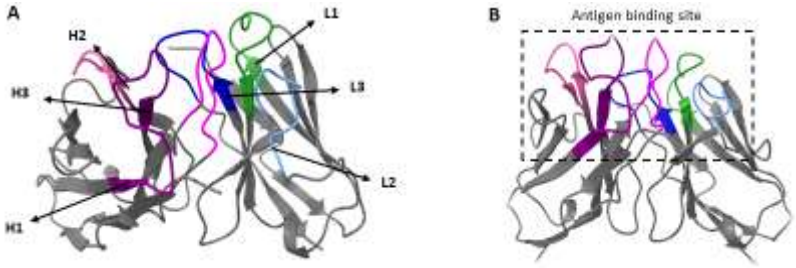

Figure 2. CDRs region on human IgG antibodies.

(A: The arrangement of the CDRs in the variable region of an antibody. The CDR is composed of three loops from the variable chains of both the heavy and light chains. These six CDRs (H1, H2, H3, L1, L2, L3) are responsible for interacting with antigens and determining the specificity and titer of the antibody; B: The antigen-binding region is formed by CDRs. These images were created using UCSF Chimera X software [17]).

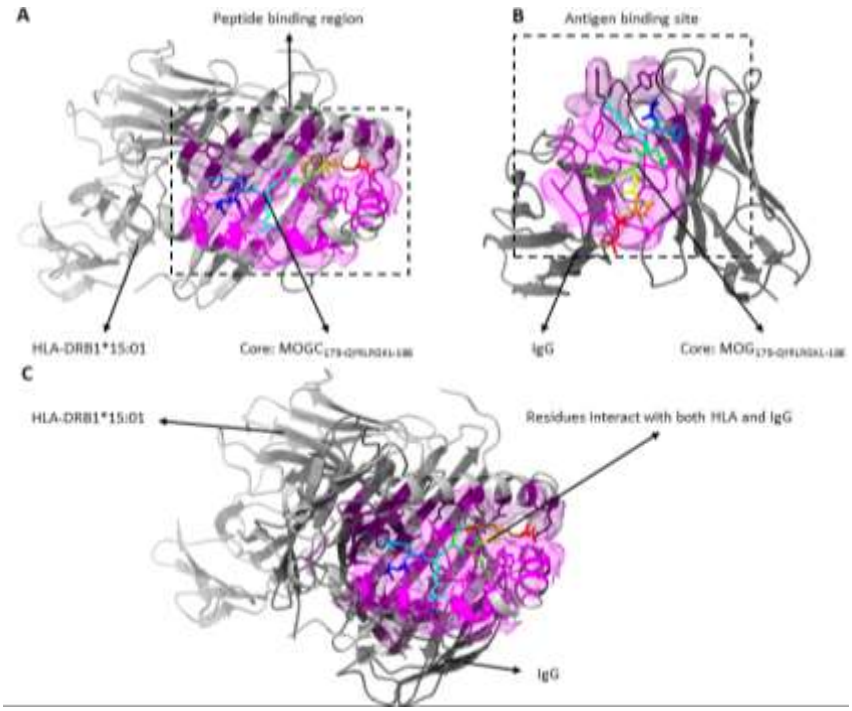

Figure 3. Epitope–HLA/IgG binding visualization.
(A: Epitope–HLA complex showing interaction sites; B: Epitope–IgG complex highlighting binding regions; C: Superposition. The region of epitope interaction on the β chain (for HLA molecules) or heavy chain (for IgG molecules) is represented in purple, while the region of epitope interaction on the α chain (for HLA molecules) or light chain (for IgG molecules) is displayed in pink. HLA is represented in dark grey and IgG in dim grey. The docking model was generated through simulation using HDOCK software, and the resulting image was displayed using UCSF Chimera X software [17]).

Table 2. Model scores of docking epitope-HLA and epitope-Fab

Ligand	$MOG_{22\text{-}LQVSSSY\text{-}28}$	$MOG_{80\text{-}YRNGKD\text{-}85}$	$MOG_{179\text{-}QYRLRGKL\text{-}186}$	$AQP4_{75\text{-}CTRKIS\text{-}80}$
Ligand	HLA-DRB1*15:01			
Docking score	-166.6	-189.61	-194.99	-144.87
Confidence score	0.5822	0.6883	0.7109	0.5744
Ligand	Fab-IgG			
Docking score	-167.74	-151.3	-187.76	-138.42
Confidence score	0.5878	0.5065	0.6803	0.5424

3.3. Similarity of Predicted Potential Core Epitopes with Peptide Sequences of Viral and Bacterial Proteins Associated to Autoimmune Diseases

To investigate the relationship between core peptides and autoimmune diseases, we analyzed them by searching the miPepBase database. Our findings revealed that the core epitope $MOG_{179\text{-}QYRLRGKL\text{-}186}$, which includes the YRLRG motif, shares similarities with three motifs previously reported in experimental data on the system. Our analysis showed a 50-71% identity, indicating a high degree of similarity to the mimotope. Furthermore, the positive values of 83-100% suggest that most of the amino acids between the core epitope and the mimotope have similar chemical properties, potentially triggering autoimmune mechanisms [19-21].

Additionally, the BLASTp tool was used to analyze the sequence similarity between the antigenic epitopes related to autoimmune diseases and all known viral and bacterial proteins in the NCBI database. The results showed that 189 proteins in 111 viral species and 145 proteins in 185 bacterial species had peptide sequences identical to 4 potential epitopes (Figure 4). Among these peptides, 37 sequences from 17 bacterial species were previously reported to be associated with autoimmune diseases. Of these, 4/37 (10.81%) peptides contained MoRF and 2/37 (5.40%) peptides contained LCR, while no sequences contained SliM. Table 3 shows some representative peptides in this group.

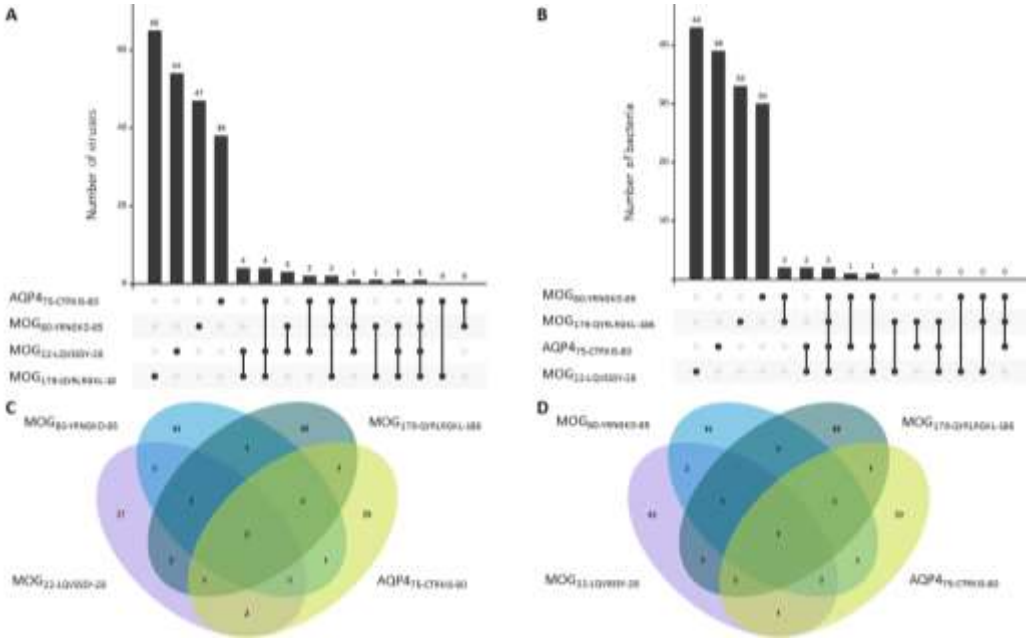

Figure 4. The number of species of viruses (A) and bacteria (B), as well as the number of virus protein sequences (C) and bacterial protein sequences (D) that contain peptide sequences identical to four potential core epitopes.
(Evenn-http://www.ehbio.com/test/venn, which was used to create the figure).

Table 3. Several peptides of viral and bacterial proteins with identity to potential epitopes in the miPepBase and NCBI databases

Sequence ID/ UniProt ID	Database name	Species/class name	Protein name	Identical peptide sequence (Position)	Identities (%)	Coverage (%)	IDPs in identical sequence	Autoimmune diseases related
$MOG_{22\text{-}LQVSSSY\text{-}28}$								
HBC95359.1	NCBI	*Clostridium sp.*	TPA: hypothetical protein	LQVSSSY (115-121)	7/7 (100%)	7/7 (100%)		MS, SLE
MBQ9637472.1	NCBI	*Prevotella sp.*	TPA: DNA-binding protein	LQVSSSY (32-38)	7/7 (100%)	7/7 (100%)	LCR	MS, SS
MDO4195505.1	NCBI	Prevotellaceae bacterium	BACON domain-containing carbohydrate-	LQVSSSY (32-38)	7/7 (100%)	7/7 (100%)		MS, SS

Sequence ID/ UniProt ID	Database name	Species/class name	Protein name	Identical peptide sequence (Position)	Identities (%)	Coverage (%)	IDPs in identical sequence	Autoimmune diseases related
			binding protein					
$MOG_{80\text{-}YRNGKD\text{-}85}$								
MDR0514596.1	NCBI	Coriobacteriaceae bacterium	leucine-rich repeat protein	YRNGKD (1364-1369)	6/6 (100%)	6/6 (100%)		SLE, SS
MCQ2219991.1	NCBI	*Prevotella sp.*	LPD5 domain-containing protein	YRNGKD (1109-1114)	6/6 (100%)	6/6 (100%)		MS, SS
$MOG_{179\text{-}QYRLRGKL\text{-}186}$								
P0AD28	miPepBase	*Escherichia coli*	Inner membrane protein	YNLRGE (556-561)	4/6 (66%)	5/6 (83%)	MoRF	EM
MBU1205559.1	NCBI	Pseudomonadota bacterium	DEAD/DEAH box helicase	QYRLRGK (70-76)	7/7 (100%)	7/7 (100%)		SLE, SS
Q4K6F8	miPepBase	*Pseudomonas fluorescens*	Lipopolysaccharide biosynthesis	EYRVVGEL (391-398)	4/8 (50%)	7/8 (87%)		MS
$AQP4_{75\text{-}CTRKIS\text{-}80}$								
MDR1028894.1	NCBI	Clostridiales Family XIII bacterium	molybdopterin-dependent oxidoreductase	CTRKIS (62-67)	6/6 (100%)	6/6 (100%)		SLE
MBS0359119.1	NCBI	Pseudomonadota bacterium	hypothetical protein	CTRKIS (343-348)	6/6 (100%)	6/6 (100%)		SLE, SS

Note: *IDPs- Intrinsically disordered proteins; MoRF-Molecular recognition features; LCRs-Low complexity regions; MS-Multiple sclerosis; SLE- Systemic lupus erythematosus; SS-Sjögren's syndrome; EM-Myalgic encephalomyelitis; The MoRF/LCRs region in the identical sequence is shown in bold*

3.4. Discussion

This study has identified four core epitopes ($MOG_{22\text{-}LQVSSSY\text{-}28}$, $MOG_{80\text{-}YRNGKD\text{-}85}$, $MOG_{179\text{-}QYRLRGKL\text{-}186}$ and $AQP4_{75\text{-}CTRKIS\text{-}80}$) with ranging 6-8 amino acids. While $CD4^+$ T-cell epitopes are generally considered to be 13-17 amino acids long, it is well-established that much shorter sequences can suffice for T-cell receptor recognition [22]. Research has demonstrated that even minimal peptide lengths of 3-5 amino acids are sufficient to activate $CD4^+$ T-cell clones specific for myelin basic protein (MBP) in the context of HLA class II molecules [23]. Borthwick et al (2020) further support this by showing that short peptides, including those 6-8 amino acids in length, can stimulate $CD4^+$ T-cell responses, leading to cytokine production such as IFN-γ and TNF-α [24]. These findings reinforce the notion that short linear epitopes, including those identified in this study, are capable of inducing $CD4^+$ T-cell mediated immune responses.

For B-cells, linear epitopes as short as 5-17 amino acids can be effectively recognized by antibodies, allowing for robust immune responses [25]. Singh et al., (2013) demonstrated that predicted B-cell epitopes were generally between 5–20 amino acids, with a significant proportion falling within the 6-8 amino acid range. These epitopes were identified based on the primary sequence of antigen and its capacity to elicit linear immune

responses [25]. A well-documented example supporting this is the ELDKWAS epitope (7 amino acids) located in the MPER domain of HIV-1 gp41, which serves as a target for the broadly neutralizing antibody 2F5. This short peptide is sufficient for efficient antibody recognition and binding, demonstrating that minimal linear epitopes are capable of inducing strong B cell-mediated immunity [26]. Collectively, this evidence confirms that short epitopes, particularly those 6-8 amino acids in length, are not only sufficient to stimulate immune responses via both $CD4^+$ T-cells and B-cells but also hold significant promise for applications in vaccine design and immunological research.

In humans, MOG_{35-55} has been identified as a predominant immunological epitope in T-cell responses in MS patients. However, no studies have demonstrated its role in MS patients and/or patients with other autoimmune diseases carrying the HLA-DRB1 15:01 allele [27]. Using ELISPOT method, Wallström et al (1998) demonstrated that $MOG_{38\text{-}GWYRPPFSRVVHL\mathbf{YRNGKD}QDGD\text{-}60}$ significantly increased the responses in number of IFN-γ-secreting cells, activated macrophages and induced antigen-specific responses in HLA-DR2 (DR*15) positive MS patients compared to controls ($P < 0,05$) [28]. In our study, two extracellular epitopes, $MOG_{22\text{-}LQVSSSY\text{-}28}$ and $MOG_{80\text{-}YRNGKD\text{-}85}$ were identified, in which $MOG_{22\text{-}LQVSSSY\text{-}28}$ has not been previously reported in other studies, while $MOG_{80\text{-}YRNGKD\text{-}85}$ is a part of $MOG_{38\text{-}GWYRPPFSRVVHL\mathbf{YRNGKD}QDGD\text{-}60}$ as published by Wallström et al. (1998) [28].

According to Weissert et al., (2002), when analyzing the T-cell response pattern of relapsing or progressive MS patients stratified by the genetic risk factor HLA-DRB1*15:01, $MOG_{145\text{-}VFLCL\mathbf{QYRLRGKL}RAE\text{-}160}$ predominantly recognized by $CD4^+$ T-cells in the intracellular part of MOG, especially the peptide fragment 146-154 in MS patients compared with the control group [29]. The peptide $MOG_{179\text{-}\mathbf{QYRLRGKL}\text{-}186}$ of our finding is a part of the peptide $MOG_{145\text{-}VFLCL\mathbf{QYRLRGKL}RAE\text{-}160}$ (bold part), suggesting that $MOG_{179\text{-}\mathbf{QYRLRGKL}\text{-}186}$ may be the core sequence of epitope. This could be a potential epitope that should be experimentally determined, particularly in relation to HLA-DRB1*15:01.

Our study identified 3 epitopes of the MOG molecule predicted to bind to both T-cells and B-cells, potentially serving as signaling epitopes in autoimmune diseases. A strong B-cell response to MOG was observed in all transgenic mice expressing HLA-DR2. The major human MOG epitopes recognized by B-cells are peptides $MOG_{1\text{-}GQFRVIGPRHPIRALVGDEVELPCRISPGK\text{-}30}$, $MOG_{51\text{-}\mathbf{YRNGKD}QDGDQAPEYRGRTELLKDAIGEGK\text{-}80}$ and $MOG_{38\text{-}GWYRPPFSRVVHL\mathbf{YRNGKD}QDGD\text{-}60}$, which have been shown to be immunodominant to T-cells in MS individuals carrying the HLA-DR2 (DRB1*15:01) allele. This demonstrates that the epitope $MOG_{80\text{-}\mathbf{YRNGKD}\text{-}85}$ in our study reacts not only with T-cells, but also with B-cells. The complex interaction between this epitope and cancer cells may play a role in determining the severity of the disease [3, 4]. Additionally, we have not found any reports to date regarding peptide fragments containing the epitopes $MOG_{22\text{-}LQVSSSY\text{-}28}$ and $MOG_{179\text{-}QYRLRGKL\text{-}186}$ that strongly interact with B-cells in patients with demyelinating diseases of the spinal cord carrying the HLA-DRB1*15:01 allele.

In addition to the MOG epitopes, our study also identified one epitope, $AQP4_{75\text{-}CTRKIS\text{-}80}$, that is shared between $CD4^+$ T-cells and B-cells. The epitopes $AQP4_{11\text{-}30}$, $AQP4_{61\text{-}80}$, $AQP4_{63\text{-}76}$, $AQP4_{91\text{-}ISGGHINPAVTVAMV\mathbf{CTRKI}\text{-}110}$, $AQP4_{139\text{-}153}$ and $AQP4_{281\text{-}305}$ were found in patients with NMOSD, MS, ADEM, and healthy subjects [30]. Strong T-cell response to $AQP4_{91\text{–}ISGGHINPAVTVAMV\mathbf{CTRKI}\text{-}110}$ in NMOSD patients has also been previously reported [31]. However, the role of $AQP4_{75\text{-}CTRKIS\text{-}80}$ in B-cells has not yet been identified.

In general, the core epitopes identified in this study are significantly shorter than previously investigated full-length peptide sequences. However, whether these truncated sequences are sufficient to elicit a robust immune response, or whether additional flanking sequences are required, remains to be

experimentally validated. Therefore, further experimental studies are needed to confirm the reliability of the *in silico* predictions presented here. Experimental approaches such as ELISPOT assays to measure T-cell responses, flow cytometry to assess activation markers, and peptide-based immunization in animal models may provide direct evidence of the immunogenic potential of these core epitopes and determine their capacity to induce both cellular and humoral immunity.

After identifying three potential core epitopes of MOG and one potential core epitope of AQP4 shared by $CD4^+$ T-cells and B-cells, the association between these autoantigenic epitopes and identical peptide sequences of viral and bacterial proteins were analyzed to provide basic information in uncovering the mechanism of disease initiation. The analysis focused on MoRFs, SLiMs, and LCRs in identical peptide sequences of published bacterial proteins associated with autoimmune diseases. Experimental results have demonstrated that the epitope $MOG_{179\text{-}QYRLRGKL\text{-}186}$ bears a striking similarity to the mimotope found in miPepBase. This suggests that the epitope has the potential to mimic the mimotopes identified in various pathologies, such as Thrombocytopenia, Encephalomyelitis, and Multiple Sclerosis. Furthermore, it is likely to elicit cross-reactive immune responses, which may lead to the development of autoimmune diseases such as MS, ADEM, and NMO [19-21]. These findings strongly support the hypothesis of molecular mimicry between the epitope and mimotopes found in viruses and bacteria associated with autoimmune diseases.

When expanding the observational database to include miPepbase and NCBI, the results showed that these regions were present at a low frequency (16.22%). MoRFs, SLiMs, and LCRs are often involved in cell signaling and protein-protein interaction networks. The replacement of microbial proteins in the host protein-protein interaction network with these regions may facilitate microbial survival and modulate host defense mechanisms [7]. However, the results of this study indicate that only a few bacteria associated with autoimmune diseases harbor IDPRs, suggesting that most bacteria with peptide sequences identical to epitopes found in autoimmune diseases may not modulate host protein functions.

It is crucial to acknowledge that the conclusions of this study are derived solely from in silico analyses, meaning that the results are predictive and potentially constrained by the accuracy and limitations of the databases, algorithms, and software tools employed. The study's findings suggest that only a small subset of bacteria associated with autoimmune diseases contain IDPRs, indicating that the majority of bacterial species with peptide sequences identical to autoimmune-associated epitopes may not actively modulate host protein functions or directly influence autoimmunity. This highlights the necessity for comprehensive experimental studies to validate the in silico predictions and elucidate the functional roles of bacterial IDPRs. Future research should incorporate structural and functional assays, mutagenesis approaches, and host-pathogen interaction models to thoroughly investigate the potential of these bacterial IDPRs to alter host immunity and contribute to the development of autoimmune diseases.

4. Conclusions

This is the first study to provide information on the HLA-DRB1*15:01 restricted T-cell and B-cell epitopes of the autoantigens MOG and AQP4 as well as MoRFs, SLiMs, and LCRs present in bacterial peptides identical to four core epitopes associated with autoimmune diseases. The discovery of four core peptides shared by both cell types will facilitate the identification of specific drugs and therapies that can target HLA-DRB1*15:01, which is thought to be susceptible to autoimmune diseases in general, and autoimmune neuropathy in particular. Additionally, this study also supports and provides information to further understand the pathogenesis of autoimmune diseases associated with MOG and AQP4.

Acknowledgements

This study was funded by project QG.22.01- Vietnam National University, Hanoi, VNU.

References

[1] A. Davidson, B. Diamond, Autoimmune Diseases, The New England Journal of Medicine, Vol. 345, No. 5, 2001, pp. 340-350, https://doi.org/10.1056/NEJM 200108023450506.

[2] D. Whittam, M. Wilson, S. Hamid, G. Keir, M. Bhojak, A. Jacob, What's New in Neuromyelitis Optica? A Short Review for the Clinical Neurologist, Journal of Neurology, Vol. 264, No. 11, 2017, pp. 2330-2344, https://doi.org/10.1007/s00415-017-8445-8.

[3] D. G. Brum, A. A. Barreira, A. C. dos Santos, D. R. Kaimen-Maciel, M. Matiello, R. M. Costa, N. H. Deghaide, L. S. Costa, P.L ouzada-Junior, P. R. Diniz, E. R. Comini-Frota, C. T. Mendes-Junior, E. A. Donadi, HLA-DRB Association in Neuromyelitis Optica is Different from that Observed in Multiple Sclerosis, Multiple Sclerosis (Houndmills, Basingstoke, England), Vol. 16, No. 1, 2010, pp. 21–29, https://doi.org/10.1177/1352458509350741.

[4] M. Matiello, J. Schaefer-Klein, D. G. Brum, E. J. Atkinson, O. H. Kantarci, B. G. Weinshenker, NMO genetics collaborators, HLA-DRB1*1501 Tagging rs3135388 Polymorphism is not Associated with Neuromyelitis Optica, Multiple Sclerosis (Houndmills, Basingstoke, England), Vol. 16, No. 8, 2010, pp. 981-984, https://doi.org/10.1177/1352458510 374340.

[5] R. Kogay, C. Schönbach, Epitope Predictions, in Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019, pp. 952-971.

[6] M. Rojas, P. R. Jiménez, D. M. Monsalve, Y. Pacheco, Y. A. Ampudia, C. R. Santana, P. S. C. Leung, A. A. Ansari, M. E. Gershwin, J. M. Anaya, Molecular Mimicry and Autoimmunity, Journal of Autoimmunity, Vol. 95, 2018, pp. 100-123, https://doi.org/10.1016/j.jaut.2018. 10.012.

[7] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, M. M. Babu, Classification of Intrinsically Disordered Regions and Proteins, Chemical Reviews, Vol. 114, No.13, 2014, pp. 6589-6631, https://doi.org/10.1021/ cr400525m.

[8] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, V. N. Uversky, Analysis of Molecular Recognition Features (MoRFs), Journal of Molecular Biology, Vol. 362, No. 5, 2006, pp. 1043-1059, https://doi.org/10.1016/ j.jmb.2006.07.087.

[9] N. E. Davey, G. Travé, T. J. Gibson, How Viruses Hijack Cell Regulation, Trends in Biochemical Sciences, Vol. 36, No. 3, 2011, pp. 159-169, https://doi.org/10.1016/j.tibs.2010.10.002.

[10] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, D. Hassabis, Highly Accurate Protein Structure Prediction with AlphaFold, Nature, Vol. 596, No. 7873, 2021, pp. 583-589, https://doi.org/10.1038/s41586-021-03828-1.

[11] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank, Nucleic Acids Research, Vol. 28, No. 1, 2000, pp. 235-242, https://doi.org/10.1093/nar/28.1.235.

[12] P. Waters, M. Woodhall, K. C. O'Connor, M. Reindl, B. Lang, D. K. Sato, M. Juryńczyk, G. Tackley, J. Rocha, T. Takahashi, T., Misu, I. Nakashima, J. Palace, K. Fujihara, M. I. Leite, A. Vincent, MOG Cell-based Assay Detects Non-MS Patients with Inflammatory Neurologic Disease, Neurology(R) Neuroimmunology & Neuroinflammation, Vol. 2, No. 3, 2015, pp. e89, https://doi.org/10.1212/NXI.0000000000000089.

[13] M. Reindl, P. Waters, Myelin Oligodendrocyte Glycoprotein Antibodies in Neurological Disease, Nature reviews. Neurology, Vol. 15, No. 2, 2019, pp. 89-102, https://doi.org/10.1038/s41582-018-0112-x.

[14] O. Schmetzer, E. Lakin, B. Roediger, A. Duchow, S. Asseyer, F. Paul, N. Siebert, Anti-aquaporin 4 IgG is Not Associated with Any Clinical Disease Characteristics in Neuromyelitis Optica Spectrum Disorder, Frontiers in Neurology, Vol. 12, No. March, 2021, pp. 1-14, https://doi.org/10.3389/fneur.2021.635419.

[15] P. Saikali, R. Cayrol, and T. Vincent, Anti-aquaporin-4 Auto-antibodies Orchestrate the Pathogenesis in Neuromyelitis Optica, Autoimmunity Reviews, Vol. 9, No. 2, 2009, pp. 132-135, https://doi.org/10.1016/j.autrev.2009.04.004.

[16] C. A. Zhang, Z. Prachanronarong, K. Gellatly, K. Marasco, W.A. Schiffer, Antibody Fab Fragment, Worldwide PDB Protein Data Bank, Vol. 01, No. 2010, 2017, pp. 1-20, https://doi.org/10.2210/pdb5JQD/pdb.

[17] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis, Protein Science, Vol. 27, 2018, pp. 14-25, https://doi.org/10. 1002/pro.3235.

[18] J. Neefjes, M. L. M. Jongsma, P. Paul, O. Bakke, Towards a Systems Understanding of MHC Class I and MHC Class II Antigen Presentation, Nature Reviews. Immunology, Vol. 11, No. 12, 2011, pp. 823-836, https://doi.org/10.1038/nri3084.

[19] Y. C. Chuang, Y. S. Lin, H. S. Liu, T. M. Yeh, Molecular Mimicry between Dengue Virus and Coagulation Factors Induces Antibodies to Inhibit Thrombin Activity and Enhance Fibrinolysis, Journal of Virology, Vol. 88, No. 23, 2014, pp. 13759-68, https://doi.org/10.1128/JVI.02166-14.

[20] J. L. Grogan, A. Kramer, A. Nogai, L. Dong, M. Ohde, J. S. Mergener, T. Kamradt, Cross-reactivity of Myelin Basic Protein-specific T Cells with Multiple Microbial Peptides: Experimental Autoimmune Encephalomyelitis Induction in TCR Transgenic Mice, Journal of Immunology, Vol. 163, No. 7, 1999, pp. 3764-3770.

[21] F. C. Westall, Molecular Mimicry Revisited: Gut Bacteria and Multiple Sclerosis, Journal of Clinical Microbiology, Vol. 44, No. 6, 2006, pp. 2099-2104, https://doi.org/10.1128/JCM.02532-05.

[22] J. S. Blum, P. A. Wearsch, P. Cresswell, Pathways of Antigen Processing, Annual Review of Immunology, No. 31, 2013, pp. 443-473, https://doi.org/10.1146/annurev-immunol-032712-095910.

[23] B. Hemmer, T. Kondo, B. Gran, C. Pinilla, I. Cortese, J. Pascal, A. Tzou, H. F. McFarland, R. Houghten, R. Martin, Minimal Peptide Length Requirements for CD4(+) T Cell Clones--implications for Molecular Mimicry and T Cell Survival, International Immunology, Vol. 12, No. 3, 2000, pp. 375-383, https://doi.org/10.1093/intimm/12.3.375.

[24] N. Borthwick, S. S. Arrieta, A. Llano, M. Takiguchi, C. Brander, T. Hanke, Novel Nested Peptide Epitopes Recognized by CD4 + T Cells Induced by HIV-1 Conserved-Region Vaccines, Vaccines, Vol. 8, No. 1, 2020, pp. 28, https://doi.org/10.3390/vaccines8010028.

[25] H. Singh, H. R. Ansari, G. P. S. Raghava, Improved Method for Linear B-cell Epitope Prediction Using Antigen's Primary Sequence, PLoS One, Vol. 8, No. 5, 2013, pp. 1-8, https://doi.org/10.1371/journal.pone.0062216.

[26] G. Ofek, M. Tang, A. Sambor, H. Katinger, J. R. Mascola, R. Wyatt, P. D. Kwong, Structure and Mechanistic Analysis of the Anti-Human Immunodeficiency Virus Type 1 Antibody 2F5 in Complex with Its gp41 Epitope, Journal of Virology, Vol. 78, No. 19, 2004, pp. 10724-10737, https://doi.org/10.1128/JVI.78.19.10724-10737.2004.

[27] N. K. D. Rosb, M. Hoffman, I. Mendel, I. Yust, J. Kaye, R. Bakimer, S. Flechter, O. Abramsky, R. Milo, A. Karni, A. B. Nun, Predominance of the Autoimmune Response to Myelin Oligodendrocyte Glycoprotein (MOG) in Multiple Sclerosis: Reactivity to the Extracellular Domain of MOG is Directed Against Three Main Regions, European Journal of Immunology, Vol. 27, No. 11, 1997, pp. 3059-3069, https://doi.org/10.1002/eji.1830271144.

[28] E. Wallstrom, M. Khademi, M. Andersson, R. Weissert, C. Linington, T. Olsson, Increased Reactivity to Myelin Oligodendrocyte Glycoprotein Peptides and Epitope Mapping in HLA DR2(15)+ Multiple Sclerosis, European Journal of Immunology, Vol. 28, No. 10, 1998, pp. 3329-3335, https://doi.org/10.1002/(SICI)1521-4141(199810) 28:10<3329::AID-IMMU3329>3.0.CO;2-B.

[29] R. Weissert, J. Kuhle, K. L. de Graaf, W. Wienhold, M. M. Herrmann, C. Müller, T. G. Forsthuber, K. H. Wiesmüller, A. Melms, High Immunogenicity of Intracellular Myelin

Oligodendrocyte Glycoprotein Epitopes, Journal of Immunology (Baltimore, Md. : 1950, Vol. 169, No. 1, 2002, pp. 548-556, https://doi.org/10.4049/ jimmunol.169.1.548.

[30] L. S. Hofer, M. Ramberger, V. Gredler, A. S. Pescoller, K. Rostásy, M. Sospedra, H. Hegen, T. Berger, A. Lutterotti, M. Reindl, Comparative Analysis of T-Cell Responses to Aquaporin-4 and Myelin Oligodendrocyte Glycoprotein in Inflammatory Demyelinating Central Nervous System Diseases, Frontiers in Immunology, Vol. 11, 2020, https://doi.org/10.3389/fimmu.2020.01188.

[31] N. Matsuya, M. Komori, K. Nomura, S. Nakane, T. Fukudome, H. Goto, H. Shiraishi, K. P. Wandinger, H. Matsuo, T. Kondo, Increased T-cell Immunity Against Aquaporin-4 and Proteolipid Protein in Neuromyelitis Optica, International Immunology, Vol. 23, No. 9, 2011, pp. 565-573, https:// doi.org/10.1093/intimm/dxr056.